WHEN RESPONSE VARIABILITY INCREASES NEURAL NETWORK ROBUSTNESS TO SYNAPTIC NOISE

Gleb Basalyga and Emilio Salinas

Department of Neurobiology and Anatomy
Wake Forest University School of Medicine
Winston-Salem, NC 27157-1010
E-mail: gbasalyg@wfubmc.edu, esalinas@wfubmc.edu

February 9, 2008

Preliminary version of paper
to appear in ***Neural Computation***

Abstract

Cortical sensory neurons are known to be highly variable, in the sense that responses evoked by identical stimuli often change dramatically from trial to trial. The origin of this variability is uncertain, but it is usually interpreted as detrimental noise that reduces the computational accuracy of neural circuits. Here we investigate the possibility that such response variability might, in fact, be beneficial, because it may partially compensate for a decrease in accuracy due to stochastic changes in the synaptic strengths of a network. We study the interplay between two kinds of noise, response (or neuronal) noise and synaptic noise, by analyzing their joint influence on the accuracy of neural networks trained to perform various tasks. We find an interesting, generic interaction: when fluctuations in the synaptic connections are proportional to their strengths (multiplicative noise), a certain amount of response noise in the input neurons can significantly improve network performance, compared to the same network without response noise. Performance is enhanced because response noise and multiplicative synaptic noise are in some ways equivalent. So, if the algorithm used to find the optimal synaptic weights can take into account the variability of the model neurons, it can also take into account the variability of the synapses. Thus, the connection patterns generated with response noise are typically more resistant to synaptic degradation than those obtained without response noise. As a consequence of this interplay, if multiplicative synaptic noise is present, it is better to have response noise in the network than not to have it. These results are demonstrated analytically for the most basic network consisting of two input neurons and one output neuron performing a simple classification task, but computer simulations show that the phenomenon persists in a wide range of architectures, including recurrent (attractor) networks and sensory-motor networks that perform coordinate transformations. The results suggest that response variability could play an important dynamic role in networks that continuously learn.

1 Introduction

Neuronal networks face an inescapable tradeoff between learning new associations and forgetting previously stored information. In competitive learning models, this is sometimes referred to as the stability-plasticity dilemma (Carpenter and Grossberg, 1987; Hertz et al., 1991): in terms of inputs and outputs, learning to respond to new inputs will interfere with the learned responses to familiar inputs. A particularly severe form of performance degradation is known as catastrophic interference (McCloskey and Cohen, 1989). It refers to situations in which the learning of new information causes the virtually complete loss of previously stored associations.

Biological networks must face a similar problem, because once a task has been mastered, plasticity mechanisms will inevitably produce further changes in the internal structural elements, leading to decreased performance. That is, within sub-networks that have already learned to perform a specific function, synaptic plasticity must at least partly appear as a source of noise. In the cortex, this problem must be quite significant, given that even primary sensory areas show a large capacity for reorganization (Wang et al., 1995; Kilgard and Merzenich, 1998; Crist et al., 2001). Some mechanisms, such as homeostatic regulation (Turrigiano and Nelson, 2000) and specific types of synaptic modification rules (Hopfield and Brody, 2004), may help alleviate the problem, but by and large, how nervous systems cope with it remains unknown.

Another factor that is typically considered as a limitation for neural computation capacity is response variability. The activity of cortical neurons is highly variable, as measured either by the temporal structure of spike trains produced during constant stimulation conditions, or by spike counts collected in a given time interval and compared across identical behavioral trials (Dean, 1981; Softky and Koch, 1992, 1993; Holt et al., 1996). Some of the biophysical factors that give rise to this variability, such as the balance between excitation and inhibition, have been identified (Softky and Koch, 1993; Shadlen and Newsome, 1994; Stevens and Zador, 1998). But its functional significance, if any, is not understood.

Here we consider a possible relationship between the two sources of randomness just discussed, whereby response variability helps counteract the destabilizing effects of synaptic changes. Although noise generally hampers performance, recent studies have shown that in nonlinear dynamical systems such as neural networks this is not always the case. The best known example is stochastic resonance, in which noise enhances the sensitivity of sensory neurons to weak periodic signals (Levin and Miller, 1996; Gammaitoni et al., 1998; Nozaki et al., 1999), but noise may play other constructive roles as well. For instance, when a system has an internal source of noise, an externally added noise can reduce the total noise of the output (Vilar and Rubi, 2000). Also, adding noise to the synaptic connections of a network during learning produces networks that, after training, are more robust to synaptic corruption and have a higher capacity to generalize (Murray and Edwards, 1994).

In this paper we study another beneficial effect of noise on neural network performance. In this case, adding randomness to the neural responses reduces the impact of fluctuations in synaptic strength. That is, here, performance depends on two sources of variability, response noise and synaptic noise, and adding some amount of response noise produces better performance than having synaptic noise alone. The reason for this paradoxical effect is that response noise acts as a regularization factor that favors connectivity matrices with many small synaptic weights over connectivity matrices with few large weights, and this minimizes the impact of a synapse that is lost or has a wrong value. We

study this regularization effect in three different cases: (1) a classification task, which in its simplest instantiation can be studied analytically, (2) a sensory-motor transformation, and (3) an attractor network that produces self-sustained activity. For the latter two, the interaction between noise terms is demonstrated by extensive numerical simulations.

2 General Framework

First we consider networks with two layers, an input layer that contains N sensory neurons and an output layer with K output neurons. A matrix $\boldsymbol{r}$ is used to denote the firing rates of the input neurons in response to M stimuli, so r_{ij} is the firing rate of input unit i when stimulus j is presented. These rates have a mean component $\overline{\boldsymbol{r}}$ plus noise, as described in detail below. The output units are driven by the first layer responses, such that the firing rate of output unit k evoked by stimulus j is

$$R_{kj} = \sum_{i=1}^{N} w_{ki}\, r_{ij}, \tag{1}$$

or in matrix form, $\boldsymbol{R} = \boldsymbol{w}\boldsymbol{r}$, where $\boldsymbol{w}$ is the $K \times N$ matrix of synaptic connections between input and output neurons. The output neurons also have a set of desired responses $\boldsymbol{F}$, where F_{kj} is the firing rate that output unit k should produce when stimulus j is presented. In other words, $\boldsymbol{F}$ contains target values that the outputs are supposed to learn. The error E is the mean squared difference between the actual driven responses R_{kj} and the desired ones,

$$E = \left\langle \frac{1}{KM} \sum_{k=1}^{K} \sum_{j=1}^{M} (R_{kj} - F_{kj})^2 \right\rangle, \tag{2}$$

or in matrix notation,

$$E = \frac{1}{KM} \left\langle \mathrm{Tr}\left[(\boldsymbol{w}\boldsymbol{r} - \boldsymbol{F})(\boldsymbol{w}\boldsymbol{r} - \boldsymbol{F})^{\mathrm{T}} \right] \right\rangle. \tag{3}$$

Here, $\mathrm{Tr}(\boldsymbol{A}) = \sum_i A_{ii}$ is the trace of a matrix and the angle brackets indicate an average over multiple trials, which corresponds to multiple samples of the noise in the inputs $\boldsymbol{r}$. The optimal synaptic connections $\overline{\boldsymbol{W}}$ are those that make the error as small as possible. These can be found by computing the derivative of Equation (3) with respect to $\boldsymbol{w}$ (or with respect to w_{ab}, if the summations are written explicitly) and setting the result equal to zero (see e.g., Golub and van Loan, 1996). These steps give

$$\overline{\boldsymbol{W}} = \boldsymbol{F}\,\overline{\boldsymbol{r}}^{\mathrm{T}} \boldsymbol{C}^{-1}, \tag{4}$$

where $\overline{\boldsymbol{r}} = \langle \boldsymbol{r} \rangle$ and $\boldsymbol{C}^{-1}$ is the inverse (or the pseudo-inverse) of the correlation matrix $\boldsymbol{C} = \left\langle \boldsymbol{r}\boldsymbol{r}^{\mathrm{T}} \right\rangle$.

The general outline of the computer experiments proceeds in five steps as follows. First, the matrix $\overline{\boldsymbol{r}}$ with the mean input responses is generated together with the desired output responses $\boldsymbol{F}$. These two quantities define the input-output transformation that the network is supposed to implement. Second, response noise is added to the mean input rates, such that

$$r_{ij} = \overline{r}_{ij}(1 + \eta_{ij}). \tag{5}$$

The random variables η_{ij} are independently drawn from a distribution with zero mean and variance σ_r^2,

$$\begin{aligned} \langle \eta_{ij} \rangle &= 0 \\ \left\langle \eta_{ij}^2 \right\rangle &= \sigma_r^2, \end{aligned} \tag{6}$$

where the brackets again denote an average over trials. We refer to this as multiplicative noise. Third, the optimal connections are found using Equation (4). Note that these connections take into account the response noise through its effect on the correlation matrix $\boldsymbol{C}$. Fourth, the connections are corrupted by multiplicative synaptic noise with variance σ_W^2, that is

$$W_{ij} = \overline{W}_{ij}(1 + \epsilon_{ij}), \tag{7}$$

where

$$\begin{aligned} \langle \epsilon_{ij} \rangle &= 0 \\ \left\langle \epsilon_{ij}^2 \right\rangle &= \sigma_W^2. \end{aligned} \tag{8}$$

Finally, the network's performance is evaluated. For this, we measure the network error E_W, which is the square error obtained with the optimal but corrupted weights $\boldsymbol{W}$, averaged over both types of noise,

$$E_W = \frac{1}{KM} \left\langle \mathrm{Tr} \left[(\boldsymbol{W}\boldsymbol{r} - \boldsymbol{F})(\boldsymbol{W}\boldsymbol{r} - \boldsymbol{F})^{\mathrm{T}} \right] \right\rangle. \tag{9}$$

Thus, the brackets in this case indicate an average over multiple trials and multiple networks, i.e., multiple corruptions of the optimal weights $\overline{\boldsymbol{W}}$.

The main result we report here is an interaction between the two types of noise: in all the network architectures that we have explored, for a fixed amount of synaptic noise σ_W, the best performance is typically found when the response noise has a certain nonzero variance. So, given that there is synaptic noise in the network, it is better to have some response noise rather than to have none.

Before addressing the first example, we should highlight some features of the chosen noise models. Regarding response noise, Equations (5, 6), other models were tested in which the fluctuations were additive rather than multiplicative. Also, Gaussian, uniform and exponential distributions were tested. The results for all combinations were qualitatively the same, so the shape of the response noise distribution does not seem to play an important role; what counts is mainly the variance. On the other hand, the benefit of response noise is observed only when the synaptic noise is multiplicative; it disappears with additive synaptic noise. However, we do test several variants of the multiplicative model, including one in which the random variables ϵ_{ij} are drawn from a Gaussian distribution and another in which they are binary, 0 or -1. The latter case represents a situation in which connections are eliminated randomly with a fixed probability.

3 Noise Interactions in a Classification Task

First we consider a task in which the two-layer, fully connected network is used to approximate a binary function. The task is to classify M stimuli on the basis of the N input firing

rates evoked by each stimulus. Only one output neuron is needed, so $K=1$. The desired response of this output neuron is the classification function

$$F_j = \begin{cases} 1 & \text{if } j \leq M/2 \\ 0 & \text{else,} \end{cases} \tag{10}$$

where j goes from 1 to M. Therefore, the job of the output unit is to produce a 1 for the first $M/2$ input stimuli and a 0 for the rest.

3.1 A Minimal Network

In order to obtain an analytical description of the noise interactions, we first consider the simplest possible network that exhibits the effect, which consists of two input neurons and two stimuli. Thus, $N=M=2$ and the desired output is $\boldsymbol{F} = (1,0)$. Note that, with a single output neuron, the matrices $\boldsymbol{W}$ and $\boldsymbol{F}$ become row vectors. Now we proceed according to the five steps outlined in the preceding section — the goal is to show analytically that, in the presence of synaptic noise, performance is typically better for a nonzero amount of response noise.

The matrix of mean input firing rates is set to

$$\overline{\boldsymbol{r}} = \begin{pmatrix} 1 & r_0 \\ r_0 & 1 \end{pmatrix}, \tag{11}$$

where r_0 is a parameter that controls the difficulty of the classification. When it is close to 1, the pairs of responses evoked by the two stimuli are very similar and large errors in the output are expected; when it is close to 0, the input responses are most different and the classification should be more accurate. After combining the mean responses with multiplicative noise, as prescribed by Equation (5), the input responses in a given trial become

$$\boldsymbol{r} = \begin{pmatrix} 1+\eta_{11} & r_0(1+\eta_{12}) \\ r_0(1+\eta_{21}) & 1+\eta_{22} \end{pmatrix}. \tag{12}$$

Assuming that the fluctuations are independent across neurons, the correlation matrix is, therefore,

$$\boldsymbol{C} = \left\langle \boldsymbol{r}\boldsymbol{r}^{\mathrm{T}} \right\rangle = \begin{pmatrix} (1+r_0^2)(1+\sigma_r^2) & 2r_0 \\ 2r_0 & (1+r_0^2)(1+\sigma_r^2) \end{pmatrix}. \tag{13}$$

Next, after calculating the inverse of $\boldsymbol{C}$, Equation (4) is used to find the optimal weights, which are

$$\begin{aligned} \overline{W}_1 &= \frac{\sigma_r^2(1+r_0^2)+(1-r_0^2)}{(1+\sigma_r^2)^2\,(1+r_0^2)^2-4r_0^2} \\ \overline{W}_2 &= \frac{\sigma_r^2(1+r_0^2)-(1-r_0^2)}{(1+\sigma_r^2)^2\,(1+r_0^2)^2-4r_0^2}\, r_0\,. \end{aligned} \tag{14}$$

Notice that these connections take into account the response variability through their dependence on σ_r. The next step is to corrupt these synaptic weights as prescribed by Equation (7), and substitute the resulting expressions into Equation (9). After making all the substitutions, calculating the averages and simplifying, we obtain the average error,

$$E_W = \frac{1}{2}\left(\sigma_W^2(\overline{W}_1^2+\overline{W}_2^2)(1+\sigma_r^2)(1+r_0^2) - \overline{W}_1 - r_0\overline{W}_2 + 1\right). \tag{15}$$

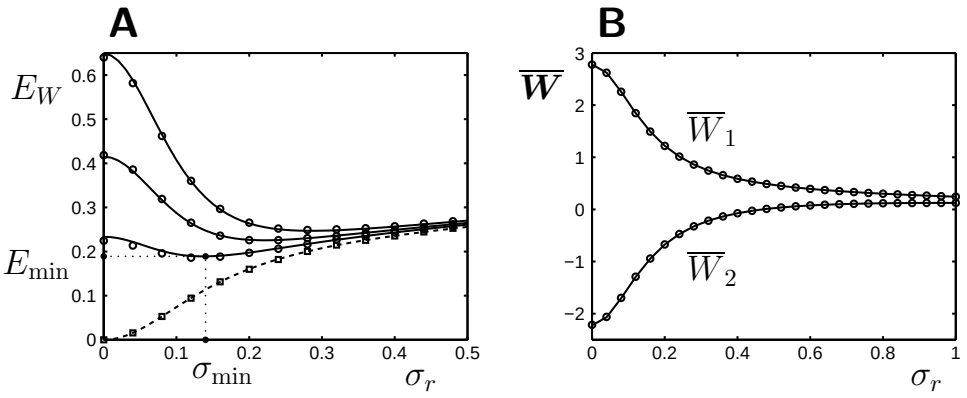

Figure 1: Noise interaction for a simple network of two input neurons and one output neuron ($K=1$, $N=M=2$). Both input responses and synaptic weights were corrupted by multiplicative Gaussian noise. For all curves, solid lines are theoretical results and symbols are simulation results averaged over 1000 networks and 100 trials per network. In all cases, $r_0=0.8$. (A) Average square difference between observed and desired output responses, E_W, as a function of the standard deviation (SD) of the response noise, σ_r. Squares and dashed line correspond to the error without synaptic noise ($\sigma_W=0$); circles and continuous lines correspond to the error with synaptic noise ($\sigma_W=0.15, 0.20, 0.25$). (B) Dependence of the (uncorrupted) optimal weights $\overline{\boldsymbol{W}}$ on σ_r.

This is the average square difference between the desired and actual responses of the output neuron given the two types of noise. It is a function only of three parameters, σ_r, σ_W and r_0, because the optimal weights themselves depend on σ_r and r_0.

The interaction between noise terms for this simple $N=K=2$ case is illustrated in Fig. 1A, which plots the error as a function of σ_r with and without synaptic variability. Here, dashed and solid lines represent the theoretical results given by Equations (14, 15) and symbols correspond to simulation results averaged over 1000 networks and 100 trials per network. Without synaptic noise (dashed line), the error increases monotonically with σ_r, as one would normally expect when adding response variability. In contrast, when $\sigma_W=0.15$, 0.2 or 0.25 (solid lines), the error initially decreases and then starts increasing again, slowly approaching the curve obtained with response noise alone.

Figure 1B shows how the optimal weights depend on σ_r. The solid lines were obtained from Equations (14) above. The curves show that the effect of response noise is to decrease the absolute values of the optimal synaptic weights. Intuitively, that is why response variability is advantageous; smaller synaptic weights also mean smaller synaptic fluctuations, because their standard deviation (SD) is proportional to the mean values. So, there is a tradeoff: the intrinsic effect of increasing σ_r is to increase the error, but with synaptic noise present, σ_r also decreases the magnitude of the weights, which lowers the impact of the synaptic fluctuations. That the impact of synaptic noise grows directly with the magnitude of the weights is also apparent from the first term in Equation (15).

The magnitude of the noise interaction can be quantified by the ratio $E_{\min}/E_0$, where the numerator is the minimal value of the error curve and the denominator is the error obtained when only synaptic noise is present, that is, when $\sigma_r=0$. The minimum error $E_{\min}$ occurs at the optimal value of σ_r, denoted as $\sigma_{\min}$. The ratio $E_{\min}/E_0$ is equal to 1 if response variability provides no advantage and approaches 0 as $\sigma_{\min}$ cancels more of the

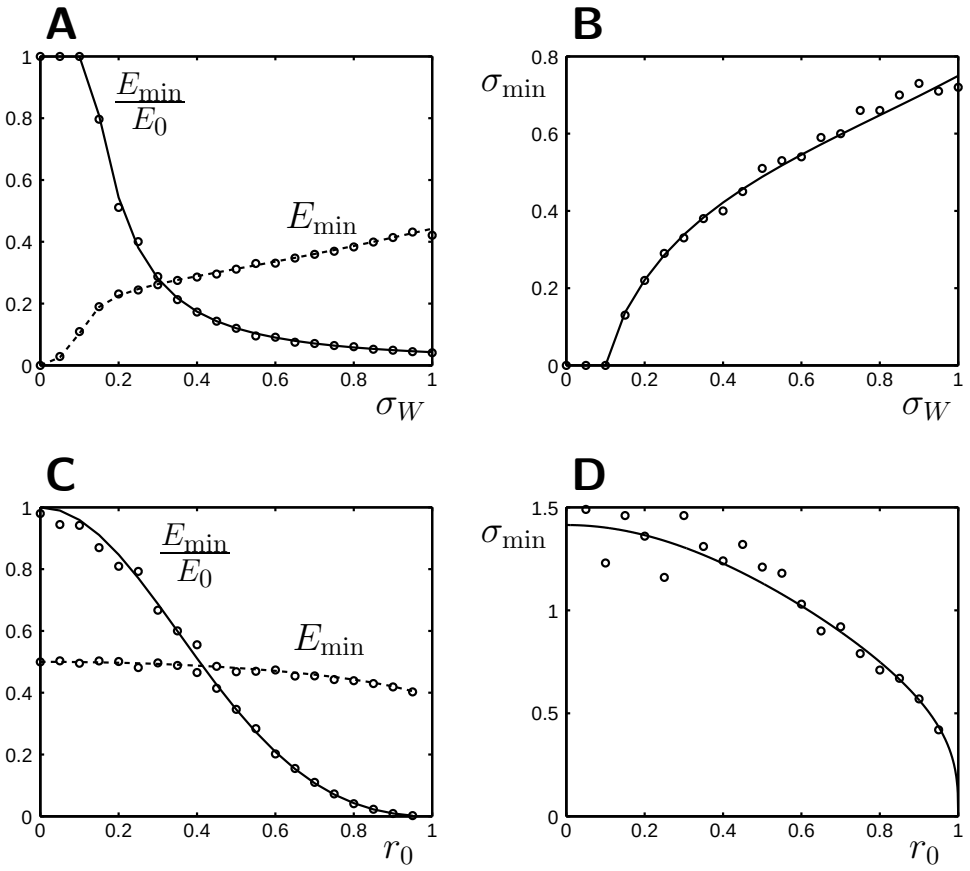

Figure 2: Optimal amount of response noise in the minimal classification network. Same network with two sensory neurons and one output neuron as in Fig. 1. Lines and symbols indicate theoretical and simulation results, respectively, averaged over 1000 networks and 100 trials per network. (A) Strength of the noise interaction quantified by $E_{\min}$ (dashed line) and $E_{\min}/E_0$ (solid line), as a function of σ_W, which determines the synaptic variability. Here and in B, $r_0 = 0.8$. (B) Optimal amount of response variability, $\sigma_{\min}$, as a function of σ_W, for the same data in A. (C) Strength of the noise interaction as a function of r_0, which parameterizes the discriminability of the mean input responses evoked by the two stimuli. Here and in D, $\sigma_W = 1$. (D) $\sigma_{\min}$, as a function of r_0 for the same data in C.

error due to synaptic noise. For the lowest solid curve in Fig. 1A the ratio is approximately 0.8, so response variability cancels about 20% of the square error generated by synaptic fluctuations. Note, however, that in these examples the error is below E_0 for a large range of values of σ_r, not only near $\sigma_{\min}$, so response noise may be beneficial even if it is not precisely matched to the amount of synaptic noise.

Figure 2 further characterizes the strength of the interaction between the two types of noise. Figures 2A, B show how the error and the optimal amount of response variability vary as functions of σ_W. These graphs indicate that the fraction of the error that σ_r is able to compensate for, as well as the optimal amount of response noise, increases with the SD of the synaptic noise. The minimum error, $E_{\min}$, grows steadily with σ_W — clearly, σ_r cannot completely compensate for synaptic corruption. Also, σ_W has to be bigger than a critical value for the noise interaction to be observed ($\sigma_W > 0.1$, approximately). However,

except when synaptic noise is very small, the optimal strategy is to add some response noise to the network.

As in the previous figure, symbols and lines in Fig. 2 correspond to simulation and theoretical results, respectively. To obtain the latter, the key is to calculate $\sigma_{\min}$. This is done by, first, substituting the optimal synaptic weights of Equation (14) into the expression for the average error, Equation (15), and second, calculating the derivative of the error with respect to σ_r^2 and equating it to zero. The resulting expression gives $\sigma_{\min}^2$ as a function of the only two remaining parameters, σ_W and r_0. The dependence, however, is highly nonlinear, so in general the solution is implicit:

$$\begin{aligned}\sigma_r^8\,(1-\sigma_W^2) + 2\sigma_r^6\,(1+a^2(1-2\sigma_W^2)) + 6\sigma_r^4 a^2\,(1-\sigma_W^2) + \\ 2\sigma_r^2 a^2\,(1+a^2+2a^2\sigma_W^2-4\sigma_W^2) + a^4(1+3\sigma_W^2) - 4a^2\sigma_W^2 \;&=\; 0\,,\end{aligned} \tag{16}$$

where

$$a \equiv \frac{1-r_0^2}{1+r_0^2}\,. \tag{17}$$

The value of σ_r that makes Equation (16) true is $\sigma_{\min}$. For Figs. 2A, B, the zero of the polynomial was found numerically for each combination of r_0 and σ_W.

Figures 2C, D show how $E_{\min}$, $E_{\min}/E_0$ and $\sigma_{\min}$ depend on the separation between evoked input responses, as parameterized by r_0. For these two plots, we chose a special case in which $\sigma_{\min}$ can be obtained analytically from Equation (16): $\sigma_W = 1$. In this particular case the dependence of $\sigma_{\min}$ on r_0 has a closed form,

$$\sigma_{\min}^2 = \frac{(1-r_0^2)^{2/3}}{1+r_0^2}\left((1+r_0)^{2/3} + (1-r_0)^{2/3}\right). \tag{18}$$

This function is shown in Fig. 2D. In general, the numerical simulations are in good agreement with the theory, except that the scatter in Fig. 2D tends to increase as r_0 approaches 0. This is due to a key feature of the noise interaction, which is that it depends on the overlap between input responses across stimuli. This can be seen as follows.

First, notice that in Fig. 2C the relative error approaches 1 as r_0 gets closer to 0. Thus, the noise interaction becomes weaker when there is less overlap between input responses, which is precisely what r_0 represents in Equation (11). If there is no overlap at all, the benefit of response noise vanishes. This fact explains why more than one neuron is needed to observe the noise interaction in the first place. This observation can be demonstrated analytically by setting $r_0 = 0$ in Equations (14) and (15), in which case the average square error becomes

$$E_W(r_0{=}0) = \frac{1}{2}\left(\frac{\sigma_W^2-1}{1+\sigma_r^2} + 1\right). \tag{19}$$

This result has interesting implications. If $\sigma_W^2 = 1$, response noise makes no difference, so there is no optimal value. If $\sigma_W^2 < 1$, the error increases monotonically with response noise, so the optimal value is 0. And if $\sigma_W^2 > 1$, the optimal strategy is to add as much noise as possible! In this case, the variance of the output neuron is so high that there is no hope of finding a reasonable solution; the best thing to do is set the mean weights to zero, disconnecting the output unit. Thus, without overlap, either the synaptic noise is so high that the network is effectively useless, or, if σ_W is tolerable, response noise does not improve performance. At $r_0 = 0$, the numerical solutions oscillate between these two extremes, producing an average error of 0.5 (leftmost point in Fig. 2C). In general, however,

with non-zero overlap there is a true optimal amount of response noise, and the more overlap there is, the larger its benefit, as shown in Fig. 2C.

The simulation data points in Fig. 2 were obtained using fluctuations ϵ and η in Equations (7) and (12), respectively, sampled from Gaussian distributions. The results, however, were virtually identical when the distribution functions were either uniform or exponential. Thus, as noted earlier, the exact shapes of the noise distributions do not restrict the observed effect.

3.2 Regularization by Noise

Above, we mentioned that response noise tends to decrease the absolute value of the optimal synaptic weights. Why is this? The reason is that minimization of the mean square error in the presence of response noise is mathematically equivalent to minimization of the same error without response noise but with an imposed constraint forcing the optimal weights to be small. This is as follows.

Consider Equation (4), which specifies the optimal weights in the two-layer network. Response noise enters into the expression through the correlation matrix. By separating the input responses into mean plus noise, we have

$$\begin{aligned} \boldsymbol{C} &= \left\langle (\overline{\boldsymbol{r}} + \boldsymbol{\eta})(\overline{\boldsymbol{r}} + \boldsymbol{\eta})^{\mathrm{T}} \right\rangle \\ &= \overline{\boldsymbol{r}}\,\overline{\boldsymbol{r}}^{\mathrm{T}} + \left\langle \boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}} \right\rangle \\ &= \overline{\boldsymbol{r}}\,\overline{\boldsymbol{r}}^{\mathrm{T}} + \boldsymbol{D}_\sigma \,, \end{aligned} \tag{20}$$

where we have assumed that the noise is additive and uncorrelated across neurons (additivity is considered for simplicity but is not necessary). This results in the diagonal matrix $\boldsymbol{D}_\sigma$ containing the variances of individual units, such that element j along the diagonal is the total variance, summed over all stimuli, of input neuron j. Thus, uncorrelated response noise adds a diagonal matrix to the correlation between average responses. In that case, Equation (4) can be rewritten as

$$\overline{\boldsymbol{W}} = \boldsymbol{F}\,\overline{\boldsymbol{r}}^{\mathrm{T}} \left(\overline{\boldsymbol{r}}\,\overline{\boldsymbol{r}}^{\mathrm{T}} + \boldsymbol{D}_\sigma \right)^{-1} . \tag{21}$$

Now consider the mean square error without any noise but with an additional term that penalizes large weights. To restrict, for instance, the total synaptic weight provided by each input neuron, add the penalty term

$$\frac{1}{KM} \sum_{i,j} \lambda_i \, w_{ij}^2 \tag{22}$$

to the original error expression, Equation (3). Here, λ_i determines how much input neuron i is taxed for its total synaptic weight. Rewriting this as a trace, the total error to be minimized in this case becomes

$$E = \frac{1}{KM} \left(\left\langle \mathrm{Tr} \left[(\boldsymbol{w}\overline{\boldsymbol{r}} - \boldsymbol{F})(\boldsymbol{w}\overline{\boldsymbol{r}} - \boldsymbol{F})^{\mathrm{T}} \right] \right\rangle + \mathrm{Tr} \left(\boldsymbol{w}^{\mathrm{T}} \boldsymbol{D}_\lambda \boldsymbol{w} \right) \right) . \tag{23}$$

where $\boldsymbol{D}_\lambda$ is a diagonal matrix that contains the penalty coefficients λ_i along the diagonal. The synaptic weights that minimize this error function are given by

$$\boldsymbol{F}\,\overline{\boldsymbol{r}}^{\mathrm{T}} \left(\overline{\boldsymbol{r}}\,\overline{\boldsymbol{r}}^{\mathrm{T}} + \boldsymbol{D}_\lambda \right)^{-1} . \tag{24}$$

But this solution has exactly the same form as Equation (21), which minimizes the error in the presence of response noise alone, without any other constraints. Therefore, adding response noise is equivalent to imposing a constraint on the magnitude of the synaptic weights, with more noise corresponding to smaller weights. The penalty term in Equation (22) can also be interpreted as a regularization term, which refers to a common type of constraint used to force the solution of an optimization problem to vary smoothly (Hinton, 1989; Haykin, 1999). Therefore, as has been pointed out previously (Bishop, 1995), the effect of response fluctuations can be described as regularization by noise.

In our model, we assumed that the fluctuations in synaptic connections are proportional to their size. What happens, then, is that response noise forces the optimal weights to be small, and this significantly decreases the part of the error that depends on σ_W. In this way, smaller synaptic weights — and therefore a nonzero σ_r — typically lead to smaller output errors.

Another way to look at the relationship between the two types of noise is to calculate the optimal mean synaptic weights taking the synaptic variability directly into account. For simplicity, suppose that there is no response noise. Substitute Equation (7) directly into Equation (3) and minimize with respect to $\overline{\boldsymbol{W}}$, now averaging over the synaptic fluctuations. With multiplicative noise the result is again an expression similar to Equations (21) and (24), where a correction proportional to the synaptic variance is added to the diagonal of the correlation matrix. In contrast, with additive synaptic noise the resulting optimal weights are exactly the same as without any variability, because this type of noise cannot be compensated for. Therefore, the recipe for counteracting response noise is equivalent to the recipe for counteracting multiplicative synaptic noise. An argument outlining why this is generally true is presented in the Discussion, Section 6.1.

3.3 Classification in Larger Networks

When the simple classification task is extended to larger numbers of first-layer neurons ($N > 2$) and more input stimuli to classify ($M > 2$), an important question can be studied: how does the interaction between synaptic and response noise depend on the dimensionality of the problem, that is, on N and M? To address this issue we did the following. Each entry in the $N \times M$ matrix $\overline{\boldsymbol{r}}$ of mean responses was taken from a uniform distribution between 0 and 1. The desired output still consisted of a single neuron's response given by Equation (10), as before. So, each one of the M input stimuli evoked a set of N neuronal responses, each set drawn from the same distribution, and the output neuron had to divide the M evoked firing rate patterns into two categories. The optimal amount of response noise was found, and the process was repeated for different combinations of N and M.

The results from these simulations are shown in Fig. 3. All data points were obtained with the same amount of synaptic variability, $\sigma_W = 0.5$. Each point represents an average over 1000 networks for which the optimal connections were corrupted. The amount of response noise that minimized the error, averaged over those 1000 corruption patterns, was found numerically by calculating the average error with the same mean responses and corruption patterns but different σ_r. For each combination of N and M, this resulted in $\sigma_{\min}$, which is shown in panel B. The actual average error obtained with $\sigma_r = \sigma_{\min}$ divided by the error for $\sigma_r = 0$ is shown in panel A, as in the previous figure. Interestingly, the benefit conferred by response noise depends strongly on the difference between N and M. With $M = 10$ input stimuli, the effect of response noise is maximized when $N = 10$ neurons are used to encode them (Fig. 3A); and viceversa, when there are $N = 10$ neurons

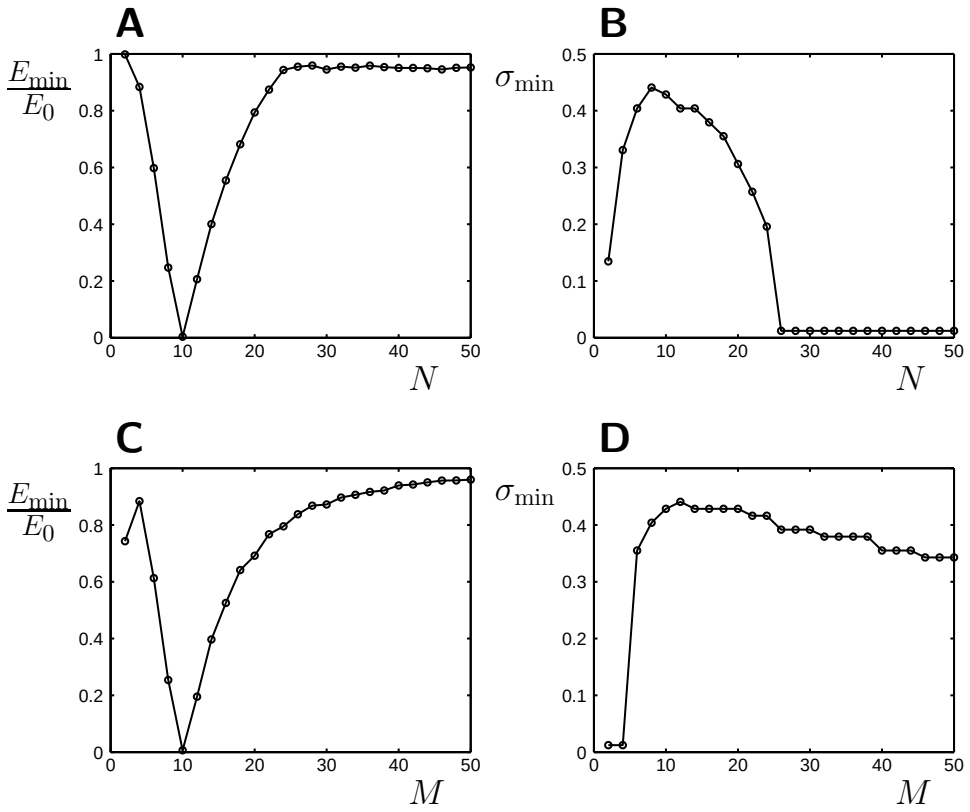

Figure 3: Interaction between synaptic noise and response noise during the classification of M input stimuli. For each stimulus, the mean responses of N input neurons were randomly selected from a uniform distribution between 0 and 1. The output unit of the network had to classify the M response patterns by producing either a 1 or a 0. The synaptic noise SD was $\sigma_W = 0.5$. Results (circles) are averages over 1000 networks and 100 trials per network. All data are from computer simulations. (A) Relative error, $E_{\min}/E_0$, as a function of the number of input neurons, N. The number of stimuli was kept constant at $M = 10$. (B) Optimal value of the response noise SD, $\sigma_{\min}$, as a function of the number of input neurons, N. Same simulations as in A. (C) Relative error as a function of the number of input stimuli, M. The number of input neurons was kept constant at $N = 10$. (D) Optimal value of the response noise SD as a function of M for the same simulations as in C.

in the network, the maximum effect is seen when they encode $M = 10$ stimuli (Fig. 3C). Results with other numbers (5, 20 and 40 stimuli or neurons) were the same: response noise always had a maximum impact when $N = M$.

This is not unreasonable. When there are many more neurons than stimuli, a moderate amount of synaptic corruption causes only a small error, because there is redundancy in the connectivity matrix. On the other hand, when there are many more input stimuli than neurons, the error is large anyway, because the N neurons cannot possibly span all the required dimensions, M. Thus, at both extremes, the impact of synaptic noise is limited. In contrast, when $N = M$ there is no redundancy but the output error can potentially be very small, so the network is most sensitive to alterations in synaptic connectivity. Thus, response noise makes a big difference when the number of responses and the number of

independent stimuli encoded are equal or nearly so. In Figs. 3A, C, the relative error is not zero for $N = M$, but it is quite small ($E_{\min} = 0.23$, $E_{\min}/E_0 = 0.004$). This is primarily because the error without any response noise, E_0, can be very large. Interestingly, the optimal amount of response noise also seems to be largest when $N = M$, as suggested by Figs. 3B, D.

In contrast to previous examples, for all data points in Fig. 3 the fluctuations in the synapses and in the firing rates, ϵ and η, were drawn from uniform rather than Gaussian distributions. As mentioned before, the variances of the underlying distributions should matter but their shapes should not. Indeed, with the same variances, results for Fig. 3 were virtually identical with Gaussian or exponential distributions.

A potential concern in this network is that, although the variability of the output neuron depends on the interaction between the two types of noise, perhaps the interaction is of little consequence with respect to actual classification performance. The relevant measure for this is the probability of correct classification, p_c. This probability is obtained by comparing the distributions of output responses to stimuli in one category versus the other, which is typically done using standard methods from signal detection theory (Dayan and Abbott, 2001). The algorithm underlying the calculation is quite simple: in each trial, the stimulus is assumed to belong to class 1 if the output firing rate is below a threshold, otherwise the stimulus belongs to class 2. To obtain p_c, the results should be averaged over trials and stimuli. Finally, note that an optimal threshold should be used to obtain the highest possible p_c. We performed this analysis on the data in Fig. 3. Indeed, p_c also depended non-monotonically on response variability. For instance, for $N = M = 10$ the values with and without response noise were $p_c(\sigma_r = \sigma_{\min}) = 0.83$ and $p_c(\sigma_r = 0) = 0.75$, where chance performance corresponds to 0.5. Also, the maximum benefit of response noise occurred for $N = M$ and decreased quickly as the difference between N and M grew, as in Figs. 3A, C. However, the amount of response noise that maximized p_c was typically about one third of the amount that minimized the mean square error. Thus, the best classification probability for $N = M = 10$ was $p_c(\sigma_r = 0.13) = 0.91$. Maximizing p_c is not equivalent to minimizing the mean square error; the two quantities weight differently the bias and variance of the output response (see Haykin, 1999). Nevertheless, response noise can also counteract part of the decrease in p_c due to synaptic noise, so its beneficial impact on classification performance is real.

4 Noise Interactions in a Sensory-Motor Network

To illustrate the interactions between synaptic and response noise in a more biologically realistic situation, we apply the general approach outlined in Section 2 to a well-known model of sensory-motor integration in the brain. We consider the classic coordinate transformation problem in which the location of an object, originally specified in retinal coordinates, becomes independent of gaze angle. This type of computation has been thoroughly studied both experimentally (Andersen et al., 1985; Brotchie et al., 1995) and theoretically (Zipser and Andersen, 1988; Salinas and Abbott, 1995; Pouget and Sejnowski, 1997), and is thought to be the basis for generating representations of object location relative to the body or the world. Also, the way in which visual and eye-position signals are integrated here is an example of what seems to be a general principle for combining different information streams in the brain (Salinas and Thier, 2000; Salinas and Sejnowski, 2001). Such integration by 'gain modulation' may have wide applicability in diverse neural cir-

cuits (Salinas, 2004), so it represents a plausible and general situation in which computational accuracy is important.

From the point of view of the phenomenon at hand, the constructive effect of response noise, this example addresses an important issue: whether the noise interaction is still observed when network performance depends on a population of output neurons. In the classification task, performance was quantified through a single neuron's response, but in this case it depends on a nonlinear combination of multiple firing rates, so maybe the impact of response noise washes out in the population average. As shown below, this is not the case.

The sensory-motor network has, as before, a feedforward architecture with two layers. The first layer contains N gain-modulated sensory units and the second or output layer contains K motor units. Each sensory neuron is connected to all output neurons through a set of feedforward connections, as illustrated in Fig. 4B. The sensory neurons are sensitive to two quantities, the location (or direction) of a target stimulus x, which is in retinal coordinates, and the gaze (or eye-position) angle y. The network is designed so that the motor layer generates or encodes a movement in a direction z, which represents the direction of the target relative to the head. The idea is that the profile of activity of the output neurons should have a single peak centered at direction z. The correct (i.e., desired) relationship between inputs and outputs is $z = x - y$, which is approximately how the angles x and y should be combined in order to generate a head-centered representation of target direction (Zipser and Andersen, 1988; Salinas and Abbott, 1995; Pouget and Sejnowski, 1997). In other words, z is the quantity encoded by the output neurons and it should relate to the quantities encoded by the sensory neurons through the function $z(x, y) = x - y$. Many other functions are possible, but as far as we can tell, the choice has little impact on the qualitative effect of response noise.

In this model, the mean firing rate of sensory neuron i is characterized by a product of two tuning functions, $f_i(x)$ and $g_i(y)$, such that

$$\overline{r}_i(x, y) = r_{\max} f_i(x) (1 - D + D g_i(y)) + r_B, \tag{25}$$

where $r_B = 4$ spikes/s is a baseline firing rate, $r_{\max} = 35$ spikes/s and D is the modulation depth, which is set to 0.9 throughout. The sensory neurons are gain modulated because they combine the information from their two inputs nonlinearly. The amplitude — but not the selectivity — of a visually-triggered response, represented by $f_i(x)$, depends on the direction of gaze (Andersen et al., 1985; Brotchie et al., 1995; Salinas and Thier, 2000). Note that, in the expression above, the second index of the mean rate $\overline{r}_{ij}$ has been replaced by parentheses indicating a dependence on x and y. This is to simplify the notation; the responses can still be arranged in a matrix $\overline{\boldsymbol{r}}$ if each value of the second index is understood to indicate a particular combination of values of x and y. For example, if the rates were evaluated in a grid with 10 x points and 10 y points, the second index would run from 1 to 100, covering all combinations. Indeed, this is how it is done in the computer.

For simplicity, the tuning curves for different neurons in a given layer are assumed to have the same shape but different preferred locations or center points, which are always between -25 and 25. Visual responses are modeled as Gaussian tuning functions of stimulus location x,

$$f_i(x) = \exp\left(-\frac{(x - a_i)^2}{2\sigma_f^2}\right), \tag{26}$$

where a_i is the preferred location and $\sigma_f = 4$ is the tuning curve width. The dependence

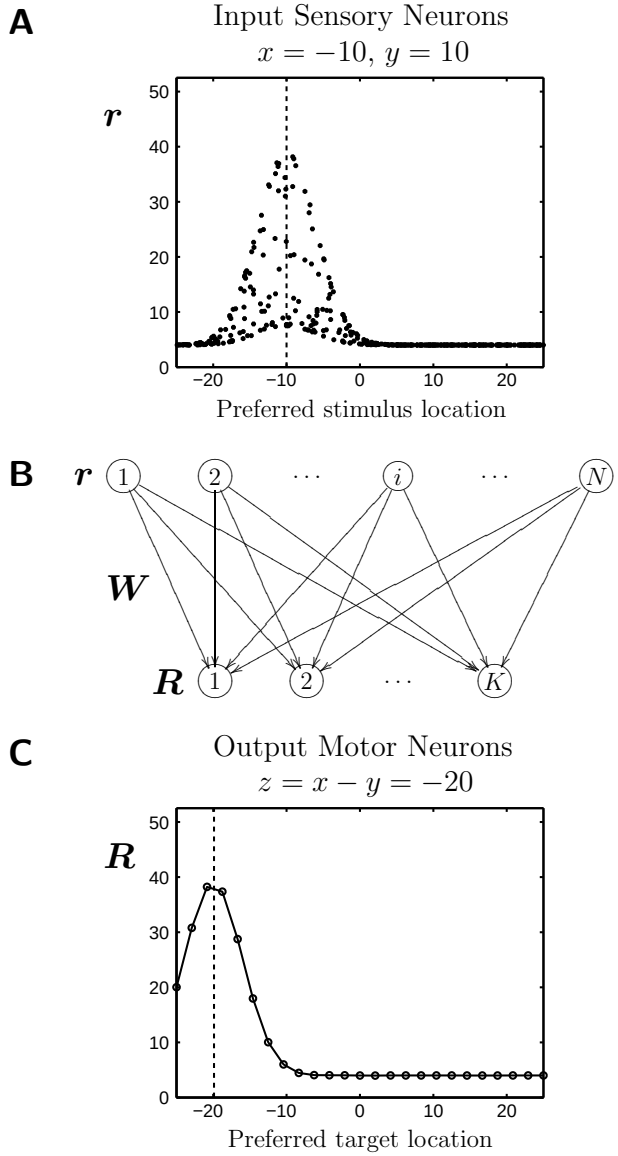

Figure 4: Network model of a sensory-motor transformation. In this network, $N = 400$, $K = 25$, $M = 400$. Target and movement directions, x and z, respectively, vary between -25 and 25, whereas gaze angle y varies between -15 and 15. The graphs correspond to a single trial in which $x = -10$, $y = 10$ and $z = x - y = -20$. Neither response noise nor synaptic corruption were included in this example. (A) Firing rates of the 400 gain-modulated input neurons arranged according to preferred stimulus location. (B) Network architecture. (C) Firing rates of the 25 output motor neurons arranged according to preferred target location.

on eye position is modeled using sigmoidal functions of the gaze angle y,

$$g_i(y) = \frac{1}{1 + \exp(-(b_i - y)/d_i)}\,, \tag{27}$$

where b_i is the center point of the sigmoid and d_i is chosen randomly between -7 and $+7$ to make sure that the curves $g_i(y)$ have different slopes for different neurons in the array.

In each trial of the task, response variability is included by applying a variant of Equation (5),

$$r_{ij} = \overline{r}_{ij} + \sqrt{\overline{r}_{ij}}\,\eta_{ij}. \tag{28}$$

This makes the variance of the rates proportional to their means, which in general is in good agreement with experimental data (Dean, 1981; Softky and Koch, 1992, 1993; Holt et al., 1996). This choice, however, is not critical (see below). The desired response for each output neuron is also described by a Gaussian,

$$F_k(z) = r_{\max}\,\exp\left(-\frac{(z-c_k)^2}{2\sigma_F^2}\right) + r_B, \tag{29}$$

where $\sigma_F = 4$ and c_k is the preferred target direction of motor neuron k. This expression gives the intended response of output unit k in terms of the encoded quantity z. Keep in mind, however, that the desired dependence on the sensory inputs is obtained by setting $z = x - y$. When driven by the first-layer neurons, the output rates are still calculated through a weighted sum,

$$R_k(z) = R_k(x,y) = \sum_{i=1}^{N} W_{ki}\, r_i(x,y). \tag{30}$$

This is equivalent to Equation (1) but with the second index defined implicitly through x and y, as mentioned above. The optimal synaptic connections $\overline{W}_{ki}$ are determined exactly as before, using Equation (4).

Typical profiles of activity for input and output neurons are shown in Figs. 4A, C for a trial with $x = -10$ and $y = 10$. The sensory neurons are arranged according to their preferred stimulus location a_i, whereas the motor neurons are arranged according to their preferred movement direction c_k. For this sample trial no variability was included; the firing rate values in Fig. 4A are scattered under a Gaussian envelope (given by Equation (26)) because the gaze-dependent gain factors vary across cells. Also, the output profile of activity is Gaussian and has a peak at the point $z = -20$, which is exactly where it should be given that the correct input-output transformation is $z = x - y$. With noise, the output responses would be scattered around the Gaussian profile and the peak would be displaced.

The error used to measure network performance is, in this case,

$$E_{\rm pop} = \langle\, |z - Z| \,\rangle\,. \tag{31}$$

This is the absolute difference, averaged over trials and networks, between the desired movement direction z — the actual head-centered target direction — and the direction Z that is encoded by the center of mass of the output activity,

$$Z = \frac{\sum_i (R_i - r_B)^2\, c_i}{\sum_k (R_k - r_B)^2}\,. \tag{32}$$

Therefore, Equation (31) gives the accuracy with which the whole motor population represents the head-centered direction of the target, whereas Equation (32) provides the recipe to read out such output activity. Now the idea is to corrupt the optimal connections and evaluate $E_{\rm pop}$ using various amounts of response noise to determine whether there is an optimum. Relative to the previous examples, the key differences are, first, that the error in

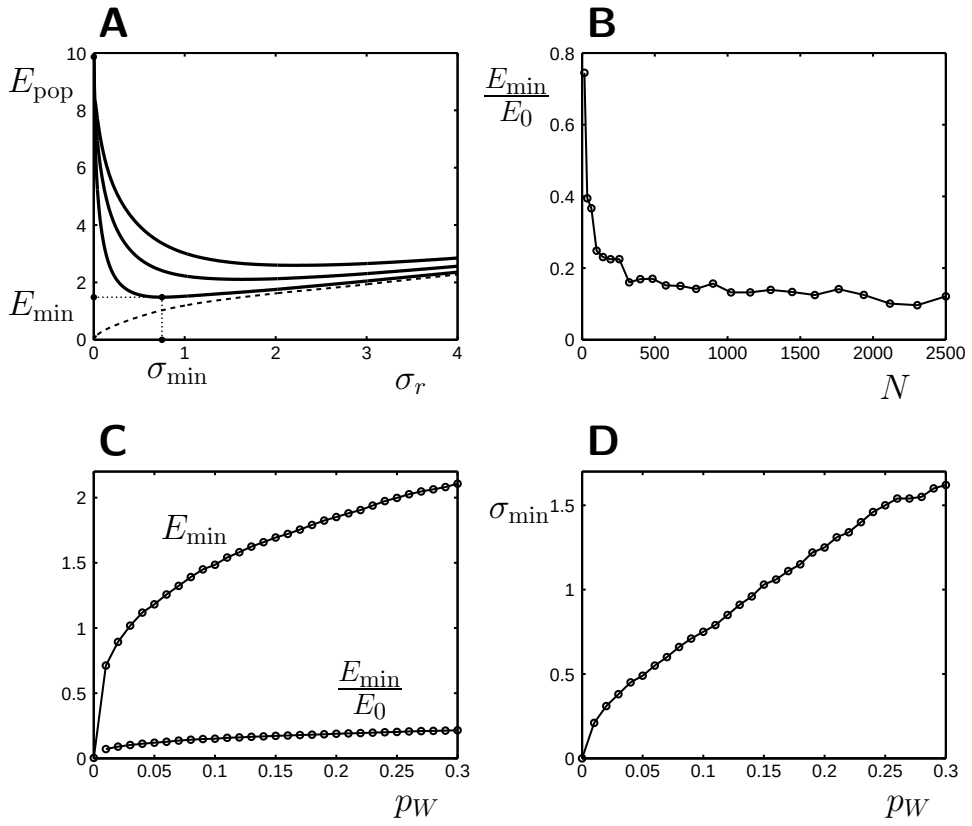

Figure 5: Noise interaction for the sensory-motor network depicted in Fig. 4. Results are averaged over 100 networks and 100 trials per network. All data are from computer simulations. (A) Average absolute deviation between actual and encoded target locations, E_{pop}, as a function of response noise. Continuous lines are for three probabilities of weight elimination, $p_W = 0.1$, 0.3 and 0.5; the dashed line corresponds to $p_W = 0$. (B) Magnitude of the noise interaction, measured by the relative error E_{min}/E_0, as a function of the number of input neurons, N, for $p_W = 0.2$. (C) E_{min} and E_{min}/E_0 as functions of p_W. (D) Optimal response noise SD, σ_{min}, as a function of p_W.

(31) represents a population average, and second, that although the connections are set to minimize the average difference between desired and driven firing rates, the performance criterion is not based directly on it.

Simulation results for this sensory-motor model are presented in Fig. 5. A total of 400 sensory and 25 output neurons were used. These units were tested with all combinations of 20 values of x and 20 values of y, uniformly spaced (thus, $M = 400$). Synaptic noise was generated by random weight elimination. This means that, after having set the connections to their optimal values given by Equation (4), each one was reset to zero with a probability p_W. Thus, on average, a fraction p_W of the weights in each network was eliminated. As shown in Fig. 5A, when $p_W > 0$, the error between the encoded and the true target direction has a minimum with respect to σ_r. These error curves represent averages over 100 networks. Interestingly, the benefit of noise does not decrease when more sensory units are included in the first layer (Fig. 5B). That is, if p_W is constant, the proportion of eliminated synapses does not change, so the error caused by synaptic corruption cannot

be reduced simply by adding more neurons.

Figure 5C shows the minimum and relative errors as functions of p_W. This graph highlights the substantial impact that response noise has on this network: the relative error stays below 0.2 even when about a third of the synapses are eliminated. This is not only because the error without response noise is high, but also because the error with an optimal amount of noise stays low. For instance, with $p_W = 0.3$ and $\sigma_r = \sigma_{\min}$, the typical deviation from the correct target direction is about 2 units, whereas with $\sigma_r = 0$ the typical deviation is about 10. Response noise thus cuts the deviation by about a factor of five, and importantly, the resulting error is still small relative to the range of values of z, which spans 50 units. Also, as observed in the classification task, in general it is better to include response noise even if σ_r is not precisely matched to the amount of synaptic variability (Fig. 5A).

Figure 5D plots $\sigma_{\min}$ as a function of the probability of synaptic elimination. The optimal amount of response noise increases with p_W and reaches fairly high levels. For instance, at a value of 1, which corresponds to p_W near 0.15, the variance of the firing rates is equal to their mean, because of Equation (28). We wondered whether the scaling law of the response noise would make any difference, so we reran the simulations with either additive noise (SD independent of mean) or noise with an SD proportional to the mean, as in Equation (5). Results in these two cases were very similar: $E_{\min}$ and $E_{\min}/E_0$ varied very much like in Fig. 5C, and the optimal amount of noise grew monotonically with p_W, as in Fig. 5D.

5 Noise Interactions in a Recurrent Network

The networks discussed in the previous sections had a feedforward architecture, and in those cases the contribution of response noise to the correlation matrix between neuronal responses could be determined analytically. In contrast, in recurrent networks the dynamics are more complex and the effects of random fluctuations more difficult to ascertain. To investigate whether response noise can still counteract some of the effects of synaptic variability, we consider a recurrent network with a well-defined function and relatively simple dynamics characterized by attractor states. When the firing rates in this network are initialized at arbitrary values, they eventually stop changing, settling down at certain steady-state points in which some neurons fire intensely and others do not. The optimal weights sought are those that allow the network to settle at predefined sets of steady-state responses, and the error is thus defined in terms of the difference between the desired steady states and the observed ones. As before, response noise is taken into account when the optimal synaptic weights are generated, although in this case the correction it introduces (relative to the noiseless case) is an approximation.

The attractor network consists of N continuous-valued neurons, each of which is connected to all other units via feedback synaptic connections (Hertz et al., 1991). With the proper connectivity, such network can generate, without any tuned input, a steady-state profile of activity with a cosine or Gaussian shape (Ben-Yishai et al., 1995; Compte et al., 2000; Salinas, 2003). Such stable 'bump'-shaped activity is observed in various neural models, including those for cortical hypercolumns (Hansel and Sompolinsky, 1998), head-direction cells (Zhang, 1996; Laing and Chow, 2001) and working memory circuits (Compte et al., 2000). Below, we find the connection matrix that allows the network to exhibit a unimodal activity profile centered at any point within the array.

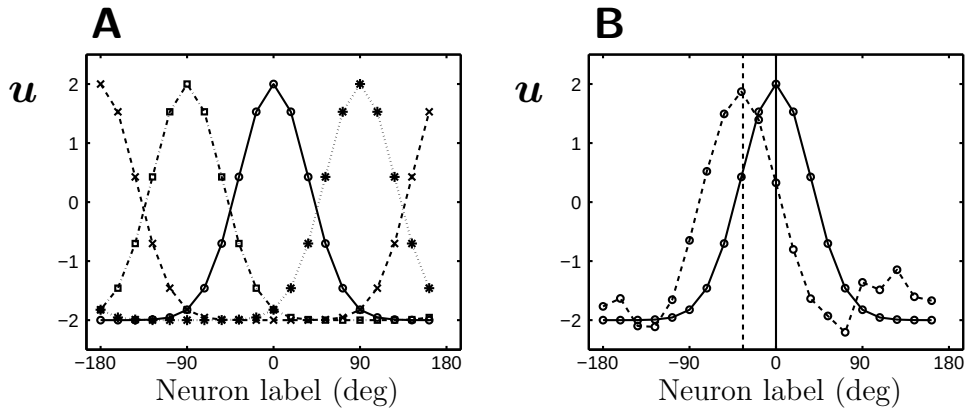

Figure 6: Steady-state responses of a recurrent neural network with 20 neurons. Results show the input currents of all units after 1000 ms of simulation time, with responses evolving according to Equation (34). Each neuron is labeled by an angle between -180° and 180°. (A) Steady-state responses for four sets of initial conditions with peaks near units -90°, 0°, +90°and 180°. The observed activity profiles are indistinguishable from the desired Gaussian curves. Neither synaptic nor response noise were included in this example. (B) Steady-state responses with and without noise. The desired activity profile is indicated by the solid line. The dotted line corresponds to the activity observed with noise after 1000 ms of simulation time, having started with an initial condition equal to the desired steady state. Vertical lines indicate the locations of the corresponding centers of mass. The absolute deviation is 34°. Here, $\sigma_r = 0.3$ and $p_W = 0.02$.

5.1 Optimal Synaptic Weights in a Recurrent Architecture

The dynamics of the network are determined by the equation

$$\tau \frac{dr_i}{dt} = -r_i + h\left(\sum_j W_{ij}\, r_j\right) + \eta_i\,, \tag{33}$$

where $\tau = 10$ is the integration time constant, r_i is the response of neuron i, and h is the activation function of the cells, which relates total current to firing rate. The sigmoid function $h(x) = 1/(1+\exp(-x))$ is used, but this choice is not critical. As before, η_i represents the response fluctuations, which are drawn independently for each neuron in every time step. In this case they are Gaussian, with zero mean and a variance $\sigma_r^2/\Delta t$. The variance of η_i is divided by the integration time step Δt to guarantee that the variance of the rate r_i remains independent of the time step (van Kampen, 1992).

For our purposes, manipulating this type of network is easier if the equations are expressed in terms of the total input currents to the cells (Hertz et al., 1991; Dayan and Abbott, 2001). If the current for neuron i is $u_i = \sum_j W_{ij}\, r_j$, then

$$\tau \frac{du_i}{dt} = -u_i + \sum_j W_{ij}\,(h(u_j) + \eta_j)\,, \tag{34}$$

is equivalent to Equation (33) above. A stationary solution of Equation (34) without input noise is such that all derivatives become zero. This corresponds to an attractor state α for

which

$$u_i^\alpha = \sum_j W_{ij}\, h(u_j^\alpha). \tag{35}$$

The label α is used because the network may have several attractors or sets of fixed points. The desired steady-state currents are denoted as U_i^α. These are Gaussian profiles of activity such that, during steady state $\alpha = 1$, neuron 1 is the most active (i.e., the Gaussian is centered at neuron 1), during steady state $\alpha = 2$, neuron 2 is the most active, and so on. Figure 6 illustrates the activity of the network at four steady states in the absence of noise ($\sigma_W = 0 = \sigma_r$). To make the network symmetric, the neurons were arranged in a ring, so their activity profiles wrap around. Because of this, each neuron is labeled with an angle. The observed currents u_i settle down at values that are almost exactly equal to the desired ones, U_i^α. The synaptic connections that achieved this match were found by enforcing the steady-state condition (35) for the desired attractors. That is, we minimized

$$E = \frac{1}{N_A} \sum_{\alpha=1}^{N_A} \sum_i \left(U_i^\alpha - \sum_j W_{ij}\, h(U_j^\alpha) \right)^2, \tag{36}$$

where U_i^α is a (wrap-around) Gaussian function of i centered at α and N_A is the number of attractors; in the simulations N_A is always equal to the number of neurons, N. This procedure leads to an expression for the optimal weights equivalent to Equation (4). Thus, without response noise,

$$\overline{\boldsymbol{W}} = \boldsymbol{L}\, \boldsymbol{C}^{-1}, \tag{37}$$

where

$$\begin{aligned} L_{ij} &= \frac{1}{N_A} \sum_\alpha U_i^\alpha\, h(U_j^\alpha) \\ C_{ij} &= \frac{1}{N_A} \sum_\alpha h(U_i^\alpha)\, h(U_j^\alpha)\,. \end{aligned} \tag{38}$$

To include the effects of response noise, we add a correction to the diagonal of the correlation matrix, as in the previous cases (see Section 3.2). We thus set

$$C_{ij} = \frac{1}{N_A} \sum_\alpha h(U_i^\alpha) h(U_j^\alpha) + \delta_{ij}\, a\, \frac{\sigma_r^2}{2\tau}, \tag{39}$$

where a is a proportionality constant. The rationale for this is as follows.

Strictly speaking, Equation (34) with response noise does not have a steady state. But consider the simpler case of a single variable u with a constant asymptotic value u_∞, such that

$$\tau \frac{du}{dt} = -u + u_\infty + \eta. \tag{40}$$

If the trajectory $u(t)$ from $t = 0$ to $t = T$ is calculated many times, starting from the same initial condition, the distribution of endpoints $u(T)$ has a well-defined mean and variance, which vary smoothly as functions of T. The mean is always equal to the endpoint that would be observed without noise, whereas for T much longer than the integration time constant τ, the variance is equal to the variance of the fluctuations on the right hand side

of Equation (40) divided by 2τ (van Kampen, 1992). These considerations suggest that we minimize

$$E = \frac{1}{N_A} \sum_{\alpha,i} \left(U_i^\alpha - \sum_j W_{ij} \left(h(U_j^\alpha) + a\,\tilde{\eta}_j \right) \right)^2 , \tag{41}$$

where the variance of $\tilde{\eta}_j$ is $\sigma_r^2/(2\tau)$. This leads to Equation (37) with the corrected correlation matrix given by (39).

5.2 Performance of the Attractor Network

To evaluate the performance of this network, we compare the center of mass of the desired activity profile to that of the observed profile tracked during a period of time. For a particular attractor α, the network is first initialized very close to that desired steady state, then Equation (34) is run for 1000 ms (100 time constants τ), and the absolute difference between the initial and the current centers of mass is recorded during the last 500 ms. The error for the recurrent networks $E_{\rm rec}$ is defined as the absolute difference averaged over this time period and all attractor states, ie., all values of α. Also, when there is synaptic noise, an additional average over networks is performed. This error function is similar to Equation (31), except that the circular topology is taken into account. Thus, $E_{\rm rec}$ is the mean absolute difference between desired and observed centers of mass. It is expressed in degrees.

Before exploring the interaction between synaptic and response noise, we used $E_{\rm rec}$ to test whether the noise-dependent correction to the correlation matrix in Equation (39) was appropriate. To do this, a recurrent network without synaptic fluctuations was simulated multiple times with different values of the parameter a and various amounts of response noise. The desired attractors were kept constant. The resulting error curves are shown in Fig. 7A. Each one gives the average absolute deviation between desired and observed centers of mass as a function of σ_r for a different value of a. The dependence on a was nonmonotonic. The optimal value we found was 0.5, which corresponds to the lowest curve (dashed) in the figure. This curve was well below the one observed without adjusting the synaptic weights. Therefore, the correction was indeed effective.

Figure 7B shows $E_{\rm rec}$ as a function of σ_r when synaptic noise is also present in the recurrent network. The three solid curves correspond to nets in which synapses were randomly eliminated with probabilities $p_W = 0.005$, 0.015 and 0.025. As with previous network architectures, a non-zero amount of response noise improves performance relative to the case where no response noise is injected. In this case, however, the mean absolute error is already about 25°at the point at which response noise starts making a difference, around $p_W = 0.005$ (Fig. 7C). This is not surprising: these types of networks are highly sensitive to changes in their synapses, so even small mismatches can lead to large errors (Seung et al., 2000; Renart et al., 2003). Also, Fig. 7C shows that the ratio $E_{\rm min}/E_0$ does not fall below 0.6, so the benefit of noise is not as large as in previous examples. The effect was somewhat weaker when synaptic variability was simulated using Gaussian noise with SD σ_W instead of random synaptic elimination. Nevertheless, it is interesting that the interaction between synaptic and response noise is observed at all under these conditions, given that the response dynamics are richer and that the minimization of Equation (41) may not be the best way to produce the desired steady-state activity.

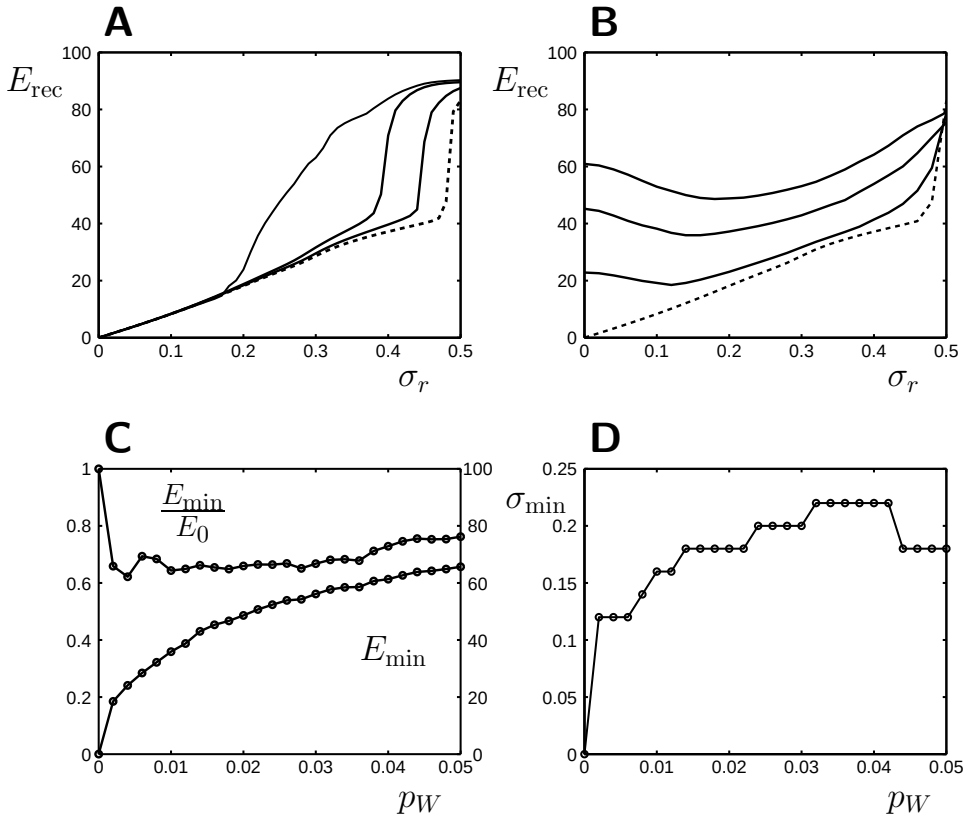

Figure 7: Interaction between synaptic and response noise in recurrent networks. (A) Average absolute difference between desired and observed centers of mass as a function of σ_r. Units are degrees. The different curves are for $a = 0$, 1.5, 1 and 0.5, from left to right. The lowest curve (dashed) was obtained with $a = 0.5$, confirming that the synaptic weights are optimized when response noise is taken into account. (B) Average error $E_{\rm rec}$ as a function of response noise. Continuous lines are for three probabilities of weight elimination $p_W = 0.005$, 0.015 and 0.025; the dashed line corresponds to $p_W = 0$. Here and in the following panels, $a = 0.5$. (C) $E_{\rm min}/E_0$ (left y-axis) and $E_{\rm min}$ (right y-axis) as functions of p_W. (D) Optimal response noise SD, $\sigma_{\rm min}$, as a function of p_W for the same data in C.

6 Discussion

6.1 Why are Synaptic and Response Fluctuations Equivalent?

We have investigated the simultaneous action of synaptic and response fluctuations on the performance of neural networks and found an interaction or equivalence between them: when synaptic noise is multiplicative, its effect is similar to that of response noise. At heart, this is a simple consequence of the product of responses and synaptic weights contained in most neural models, which has the form $\sum_j W_j r_j$. With multiplicative noise in one of the variables, this weighted sum turns into $\sum_j W_j(1 + \xi_j)r_j$, which is the same whether it is the synapse or the response that fluctuates. In either case, the total stochastic component $\sum_j W_j \xi_j r_j$ scales with the synaptic weights. The same result is obtained with additive response noise. Additive synaptic noise behaves differently, however. It instead leads to a total fluctuation $\sum_j \xi_j r_j$ that is independent of the mean weights. Evidently, in this case

the mean values of the weights have no effect on the size of the fluctuations. Thus, the key requirement for some form of equivalence between the two noise sources is that the synaptic fluctuations must depend on the strength of the synapses.

This condition was applied to the three sets of simulations presented above, which corresponded to the classification of arbitrary response patterns, a sensory-motor transformation, and the generation of multiple self-sustained activity profiles. This selection of problems was meant to illustrate the generality of the observations outlined in the above paragraph. And indeed, although the three problems differed in many respects, the results were qualitatively the same.

We should also point out that, in all the simulations, the criterion used to determine the optimality of the synaptic weights was based on a mean square error. But perhaps the noise interaction changes when a different criterion is used. To investigate this, we performed additional simulations of the small 2×1 network in which the optimal synaptic weights were those that minimized a mean absolute deviation; thus, the square in Equation (2) was substituted with an absolute value. In this case everything proceeded as before, except that the mean weight values $\overline{W}$ had to be found numerically. For this, the averages were performed explicitly and the downhill simplex method was used to search for the best weights (Press et al., 1992). The results, however, were very similar to those in Fig. 2A. Although the shapes of the curves were not exactly the same, the relative and minimum errors found with the absolute value varied very much like with the mean-square error criterion as functions of σ_W. Therefore, our conclusions do not seem to depend strongly on the specific function used to weight the errors and find the best synaptic connection values.

6.2 When Should Response Noise Increase?

According to the argument above, the most general way to state our results is this: assuming that neuronal activities are determined by weighted sums, any mechanism that is able to dampen the impact of response noise will automatically reduce the impact of multiplicative synaptic noise as well. Furthermore, we suggest that under some circumstances it is better to add more response noise and increase the dampening factor, than ignore the synaptic fluctuations altogether. There are two conditions for this scenario to make sense. (1) The network must be highly sensitive to changes in connectivity. This can be seen, for instance, in Fig. 3A, which shows that the highest benefit of response noise occurs when the number of neurons matches the number of conditions to be satisfied — it is at this point that the connections need to be most accurate. (2) The fluctuations in connectivity cannot be evaluated directly. That is, why not take into account the synaptic noise in exactly the same way as the response noise when the optimal connections are sought? For example, the average in Equation (3) could also include an average over networks (synaptic fluctuations), in which case the optimal mean weights would depend not only on σ_r but also on σ_W. In the simulations this could certainly be done, and would lead to smaller errors. But we explicitly consider the possibility that either σ_W is unknown a priori, or there is no separate biophysical mechanism for implementing the corresponding corrections to the synaptic connections.

Condition number 2 is not unreasonable. Realistic networks with high synaptic plasticity must incorporate mechanisms to ensure that ongoing learning does not disrupt their previously acquired functionality. Thus, synaptic modifications rules need to achieve two goals: to establish new associations that are relevant for the current behavioral task, and to

make adjustments to prevent interference from other, future associations. The latter may be particularly difficult to achieve if learning rates change unpredictably with time. It is not clear whether plausible (e.g., local) synaptic modification mechanisms could solve both problems simultaneously (see Hopfield and Brody, 2004), but the present results suggest an alternative: synaptic modification rules could be used exclusively to learn new associations based on current information, whereas response noise could be used to indirectly make the connectivity more robust to synaptic fluctuations. Although this mechanism evidently doesn't solve the problem of combining multiple learned associations, it might alleviate it. Its advantage is that, assuming that neural circuits have evolved to adaptively optimize their function in the face of true noise, simply increasing their response variability would generate synaptic connectivity patterns that are more resistant to fluctuations.

6.3 When is Synaptic Noise Multiplicative?

The condition that noise should be multiplicative means that changes in synaptic weight should be proportional to the magnitude of the weight. Evidently, not all types of synaptic modification processes lead to fluctuations that can be statistically modeled as multiplicative noise; for instance, saturation may prevent positive increases, thus restricting the variability of strong synapses. However, synaptic changes that generally increase with initial strength should be reasonably well approximated by the multiplicative model. Random synapse elimination fits this model because, if a weak synapse disappears, the change is small, whereas if a strong synapse disappears, the change is large. Thus, the magnitude of the changes correlates with initial strength. Another procedure that corresponds to multiplicative synaptic noise is this. Suppose the size of the synaptic changes is fixed, so that weights can only vary by $\pm\delta w$, but suppose also that the probability of suffering a change increases with initial synaptic strength. In this case, all changes are equal, but on average a population of strong synapses whould show higher variability than a population of weak ones. In simulations, the disruption caused by this type of synaptic corruption is indeed lessened by response noise (data not shown).

6.4 Final Remarks

To summarize, the scenario we envision rests on five critical assumptions: (1) the activity of each neuron depends on synaptically-weighted sums of its (noisy) inputs, (2) network performance is highly sensitive to changes in synaptic connectivity, (3) synaptic changes unrelated to a function that has already been learned can be modeled as multiplicative noise, (4) synaptic modification mechanisms are able to take into account response noise, so synaptic strengths are adjusted to minimize its impact, but (5) synaptic modification mechanisms do not directly account for future learning. Under these conditions, our results suggest that increasing the variability of neuronal responses would, on average, result in more accurate performance. Although some of these assumptions may be rather restrictive, the diversity of synaptic plasticity mechanisms together with the high response variability observed in many areas of the brain make this constructive noise effect worth considering.

Acknowledgments.

Research was supported by NIH grant NS044894.

References

Andersen, R. A., Essick, G. K., and Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230:450–458.

Ben-Yishai, R., Bar-Or, R. L., and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *PNAS*, 92:3844–3848.

Bishop, C. M. (1995). Training with noise is equivalent to tikhonov regularization. *Neural Computation*, 7:108–116.

Brotchie, P. R., Andersen, R. A., Snyder, L. H., and Goodman, S. J. (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature*, 375:232–235.

Carpenter, G. A. and Grossberg, S. (1987). Art2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, 26:4919–4930.

Compte, A., Brunel, N., Goldman-Rakic, P., and Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, 10:910–23.

Crist, R. E., Li, W., and D.Gilbert, C. (2001). Learning to see: experience and attention in primary visual cortex. *Nature Neuroscience*, 4(4):519–525.

Dayan, P. and Abbott, L. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. MIT Press.

Dean, A. (1981). The variability of discharge of simple cells in the cat striate cortex. *Exp Brain Res*, 44:437–440.

Gammaitoni, L., Hänggi, P., Jung, P., and Marchesoni, F. (1998). Stochastic resonance. *Rev. Mod. Phys.*, 70:223–287.

Golub, G. H. and van Loan, C. F. (1996). *Matrix Computations*. The John Hopkins University Press, Baltimore, 3 edition.

Hansel, D. and Sompolinsky, H. (1998). Modeling feature selectivity in local cortical circuits. In Koch, C. and Segev, I., editors, *Methods in Neuronal Modeling: From Synapse to Networks.*, pages 499–567. MIT Press, Cambridge, MA.

Haykin, S. (1999). *Neural Networks. A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice Hall.

Hertz, J., Krogh, A., and Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*. Addison-Wesley, New York.

Hinton, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence*, 40:185–234.

Holt, G. R., Softky, W. R., Koch, C., and Douglas, R. J. (1996). Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal Neurophysiology*, 75:1806–1814.

Hopfield, J. J. and Brody, C. D. (2004). Learning rules and network repair in spike-timing-based computation networks. *Proc Natl Acad Sci USA*, 101:337–342.

Kilgard, M. P. and Merzenich, M. M. (1998). Plasticity of temporal information processing in the primary auditory cortex. *Nature Neuroscience*, 1:727–731.

Laing, C. R. and Chow, C. C. (2001). Stationary bumps in networks of spiking neurons. *Neural Computation*, 13(7):1473–1494.

Levin, J. E. and Miller, J. P. (1996). Broadband neural encoding in the cricket cercal sensory system enhanced by stochastic resonance. *Nature*, 380:165–168.

McCloskey, M. and Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *The Psychology of Learning and Motivation*, 24:109–165.

Murray, A. F. and Edwards, P. J. (1994). Enhanced MLP performance and fault tolerance resulting from synaptic weight noise during training. *IEEE Transactions on Neural Networks*, 5(5):792–802.

Nozaki, D., Mar, D. J., Grigg, P., and Collins, J. J. (1999). Effects of colored noise on stochastic resonance in sensory neurons. *Physical Review Letters*, 82:2402–2405.

Pouget, A. and Sejnowski, T. J. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*, 9:222–237.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*. Cambridge University Press, New York.

Renart, A., Song, P., and Wang, X. J. (2003). Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron*, 38:473–485.

Salinas, E. (2003). Background synaptic activity as a switch between dynamical states in a network. *Neural Computation*, 15(7):1439–1475.

Salinas, E. (2004). Context-dependent selection of visuomotor maps. *BMC Neuroscience*, 5(1):47.

Salinas, E. and Abbott, L. F. (1995). Transfer of coded information from sensory to motor networks. *Journal of Neuroscience*, 15:6461–6474.

Salinas, E. and Sejnowski, T. J. (2001). Gain modulation in the central nervous system: where behavior, neurophysiology and computation meet. *Neuroscientist*, 2:539–550.

Salinas, E. and Thier, P. (2000). Gain modulation: a major computational principle of the central nervous system. *Neuron*, 27:15–21.

Seung, H. S., Lee, D. D., Reis, B. Y., and Tank, D. W. (2000). Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron*, 26:259–271.

Shadlen, M. N. and Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.*, 4:569–579.

Softky, W. P. and Koch, C. (1992). Cortical cells should fire regularly, but do not. *Neural Computation*, 4(5):643–646.

Softky, W. R. and Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *Journal of Neuroscience*, 13:334–350.

Stevens, C. F. and Zador, A. M. (1998). Input synchrony and the irregular firing of cortical neurons. *Nature Neuroscience*, 1:210–217.

Turrigiano, G. G. and Nelson, S. B. (2000). Hebb and homeostasis in neuronal plasticity. *Curr Opin Neurobiol*, 10:358–364.

van Kampen, N. G. (1992). *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam.

Vilar, J. M. G. and Rubi, J. M. (2000). Scaling of Noise and Constructive Aspects of Fluctuations. *Lecture Notes in Physics, Berlin Springer Verlag*, 557:121.

Wang, X., Merzenich, M. M., Sameshima, K., and Jenkins, W. (1995). Remodelling of hand representation in adult cortex determined by timing of tactile stimulation. *Nature*, 378:71–75.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *Journal of Neuroscience*, 16(6):2112–2126.

Zipser, D. and Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331:679–684.